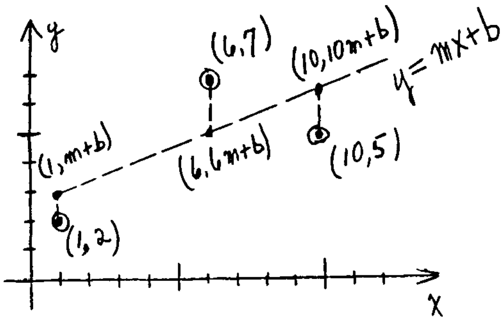
LEAST SQUARE CURVE-FITTING BY ALGEBRA METHODS

Kenneth Cummins
Kent State University
Kent, Ohio

In developing the equation of the "line of best fit" by the least squares method, one usually thinks of the use of the calculus, but it can be done by the high school method of "completing the square." For example, what are the values of m and b which make the sum of the squares of the ordinate differences from the line to the experimental points a minimum -- that is, for what values of m and b does



$$S = (m + b - 2)^2 + (6m + b - 7)^2 + (10m + b - 5)^2 \qquad (1)$$

have minimum value? The calculus uses $\frac{\partial S}{\partial m}$ and $\frac{\partial S}{\partial b}$,

but we shall first of all simplify (1) above to get

$$S = 137\,m^2 + 3b^2 + 78 + 34mb - 188m - 28b \qquad (2)$$

and write S first as a quadratic in m and then as a quadratic in b and we have

$$S = 137m^2 + (34b - 188)\,m + (3b^2 - 28b + 78) \text{ and} \qquad (3)$$
$$S = 3b^2 + (34m - 28)b + (137m^2 - 188m + 78). \qquad (4)$$

Treating further, to prepare to "complete the square" we write

$$\frac{S}{137} = m^2 + \frac{34b - 188}{137} m + \underline{\qquad\qquad} \tag{5}$$

$$\frac{S}{3} = b^2 + \frac{34m - 28}{3} b + \underline{\qquad\qquad} \tag{6}$$

Completing the square, one has

$$\frac{S}{137} = (m + \frac{17b - 94}{137})^2 + \underline{\qquad\qquad} \tag{7}$$

$$\frac{S}{3} = (b + \frac{17m - 14}{3})^2 + \underline{\qquad\qquad} \tag{8}$$

where the terms omitted in (5), (6), (7), (8) do not enter into the present discussion. For S to have minimum value, the following two relations obtained from (7) and (8) must hold simultaneously:

$$137m + 17b = 94$$
$$17m + 3b = 14.$$

We get $m = 0.4$ and $b = 2.6$ (to the nearest tenth) so

$$y = 0.4x + 2.6$$

is the "least square" line of best fit.

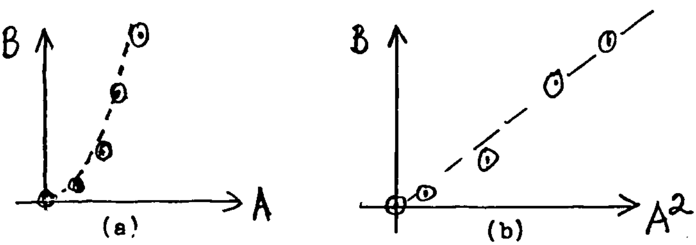
One may easily apply this method to the three points (a_1,b_1), (a_2,b_2) and (a_3,b_3) and the relations obtained are

$$m(a_1+a_2+a_3) + 3b = b_1 + b_2 + b_3$$

$$m(a_1^2 + a_2^2 + a_3^2) + (a_1 + a_2 + a_3)b = a_1b_1 + a_2b_2 + a_3b_3$$

and a generalization is suggested.

It is seen that this problem is within the grasp and capability of advanced algebra students and the results may well become a tool in the study of experimental data. Indeed, it can be extended to the study of data which seem to vary in ways other than linearly. If the original data suggest a curve as in (a) below and on squaring the "A-data" one obtains an approximate straight line as seen in (b), and if then one uses the points (A^2, B) in the "line of best fit study," the result will read $B = mA^2 + b$ and the result will be the "parabola of best fit."



Some studies may require one to cube the A-data, or to invert the numbers, or to use log A before getting a curve close to a straight line. In all cases one can make proper changes to arrive at the equation of the "curve of best fit."

In the January 1982 (Vol. 75, No. 1, pp. 57-61) issue of the Mathematics Teacher, John Staib has a well-written article on "Line Fitting Using Only High School Algebra." However, his method uses transformations and involves much more mathematics than the above.

With approaches to the problem through algebra the upper high school or beginning college student is off to exciting adventure in the study of relations. The mathematics laboratory, or experiments in the classroom, may well provide opportunity to carry on some mathematical studies done in the sciences and thus see "mathematics at work" in arriving at curve of best fit relations. -- There is really no end to this!